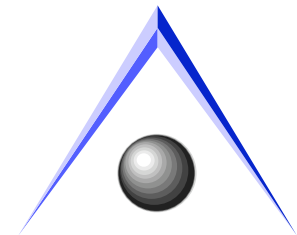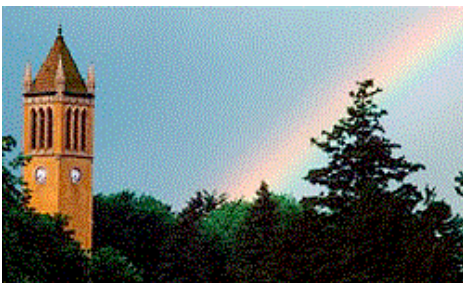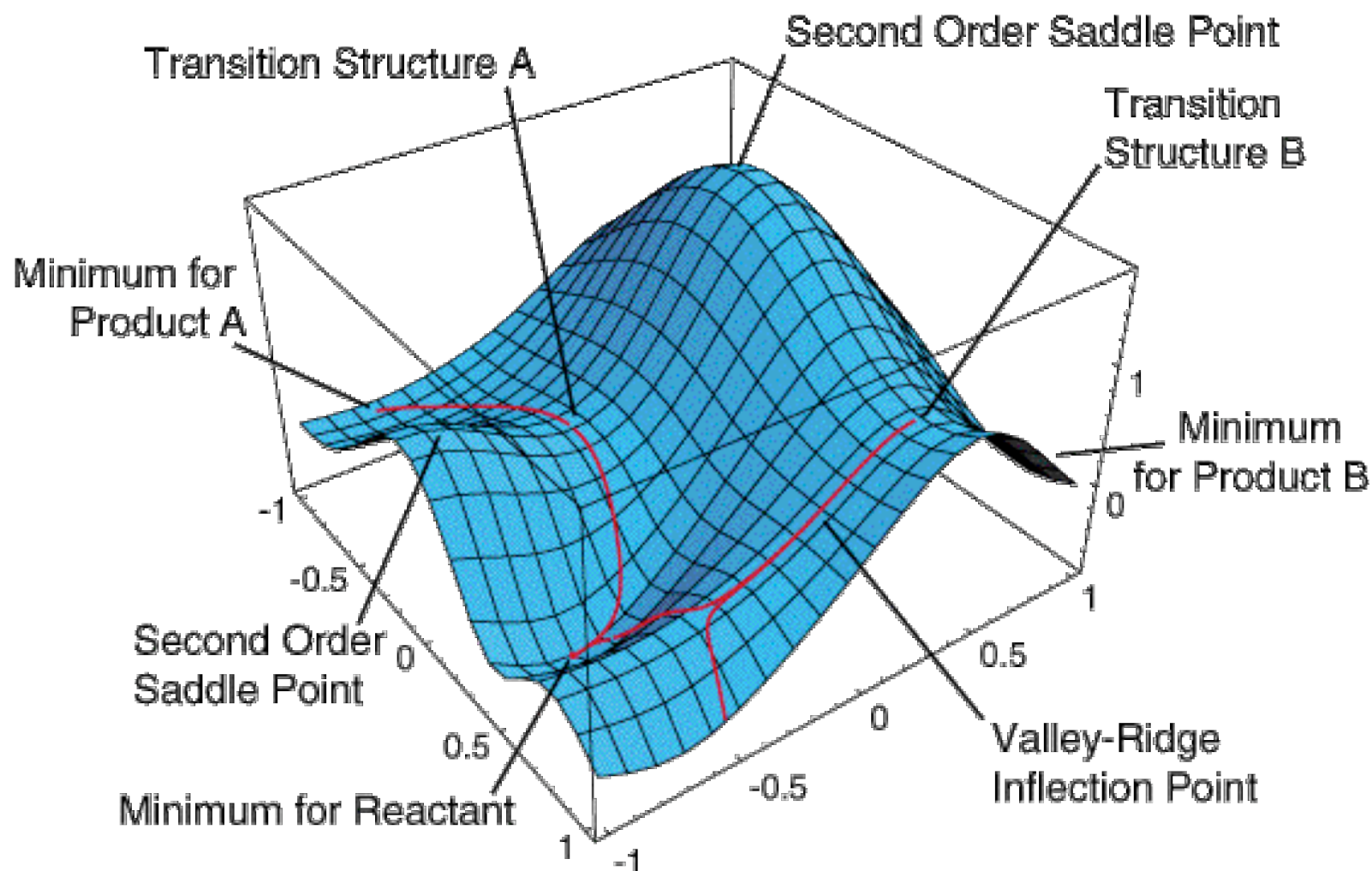# ENERGY LANDSCAPES: FINDING GLOBAL MINIMA ON COMPLEX POTENTIAL ENERGY SURFACES

Mark S. Gordon

Iowa State University

Ames Laboratory

Potential Energy Surfaces

H. B. Schlegel

NEED FOR FINDING MULTIPLE MINIMA

- Most experiments performed at finite temperatures
- Experimental observables are usually <u>Boltzmann averages</u> over large number of <u>microstates</u>
  - Large number of local minima
  - Increases rapidly as # molecules increases
- Even at 0K, molecules have zero point vibrational energy

MONTE CARLO APPROACH

- Need a method that efficiently samples <u>phase space</u>
    - 6-N dimensional space (N = # atoms)
    - 3N coordinates (x_i, y_i, z_i), momenta (p_{xi}, p_{yi}, p_{zi})
- Monte Carlo samples phase space using random numbers (hence, the name)
- Cannot be completely random
    - Would choose many chemically meaningless structures with very high energies
        - e.g. completely dissociated H_2O
        - Two superimposed water molecules

MONTE CARLO APPROACH

- Want to select important (high probability) points (structures)
  - Metropolis/importance sampling
  - Choose configurations based on Boltzmann probability
    - $P_i = \exp(-\Delta E_i / kT)$
    - ΔE_i = relative energy, k = Boltzmann const, T = temp
  - Start with "reasonable" configuration $\{q_i\}_1$
  - Perturb $\{q_i\}_1$ to new configuration $\{q_i\}_2$ by moving one or more atoms

SELECTING NEW CONFIGURATIONS

– Accept or reject new configuration $\{q_i\}_2$:

- $E_2 < E_1$: probability for acceptance=1
- $E_2 > E_1$, probability for acceptance=$\exp(-(E_2-E_1)/kT)$
  - So, very high energy configurations are rejected
  - If random # <Boltzmann factor, new configuration accepted
- Requires calculation of energy, not gradient

– General procedure

- Random # determines displacement
- Generate new configuration $\{q_2\}$
- Accept/reject $\{q_2\}$
- Can calculate designated properties (dipole moment)
- Repeat (from current configuration)

PRACTICAL CONSIDERATIONS

- If thermodynamic properties required
  - Need an equilibration step (MD)
- Step sizes:
  - Too small: too small fraction of configuration space sampled
  - Too large: get into unimportant high energy regions
    - High rejection rate
    - Leads to inefficient sampling of configuration space
- Periodic boundary conditions (PBC)
  - Keeps population of phase space constant

CONVERGENCE

- Achieved when sufficient # configuration points sampled
  - How can we tell?
  - Could get trapped in high energy state & never get out
  - Run for longer times, different starting structures
    - Not necessarily successful
  - Simulated annealing
    - Heat system to some temperature
    - Increases acceptance probability $\exp(-(E_2-E_1)/kT)$
    - Cool system slowly to trap in lower energy minimum

OTHER CONSIDERATIONS

- Generally not useful for transition states
  - High energy structures generally rejected
- Difficult to apply to large flexible species
  - Polymers, proteins
  - High energy torsional structures often rejected
  - Non-equilibrium statistical mechanics (kinetic Mone Carlo) can simulate high energy species
    - Jim Evans' Math 526

MONTE CARLO IN GAMESS

- Combined with simulated annealing
- Used to find large numbers of low-energy structures
- Currently 2 options
  - Any combination of EFPs
  - Quantum solute with fixed geometry + EFPs
    - Can fully optimize QM + EFP every n steps
- Can run using parallel computations
- RUNTYP=GLOBOP in $CONTRL

```
$contrl  scftyp=rohf runtyp=globop   coord=unique mult=2 exetyp=run $end
$system mwords=7  timlim=99999 $end
$basis   gbasis=n31 ngauss=6  npfunc=1 ndfunc=1 diffsp=.t. diffs=.t. $end
$statpt NPRT=-2 NPUN=-2  nstep=200 $end
$scf    diis=.t.  $end
$globop  tempi=5000 tempf=200 mcmin=.t. nstmin=100
   xmin=-11 xmax=12 ymin=-9.5 ymax=10 zmin=-10  zmax=10
   nfrmov=3 ntemps=200 ntran=10  ngeopt=200 optn=.t.
$end
$DATA
ab initio glycyl radical neutral with 5 efp waters: Best E=-282.320912
C1
 C          6.0   1.4366227073  -1.3618857572   4.0655573803
 O          8.0   0.4566353708  -1.8093400805   3.3158298961
 H          1.0   0.0592366109  -1.1180331412   2.7765656232
 O          8.0   1.8034700099  -0.2075867796   4.0879415156
 C          6.0   2.0448699336  -2.3803596920   4.8842698007
 H          1.0   1.7144395836  -3.3971151189   4.8446837573
 N          7.0   3.1310914629  -2.0532214656   5.6387235922
 H          1.0   3.3132974655  -1.0802147193   5.7518721353
 H          1.0   3.3583481085  -2.6303097320   6.4146879550
$END
$efrag

FRAGNAME=H2OEF2
ZO1             -1.6098833493   3.4141602757   0.2879216046
ZH2             -2.3755312746   3.0796589011  -0.1511348142
ZH3             -1.7579600714   3.3036524265   1.2135239780
```

ADDITIONAL CONSIDERATIONS

- Impact of simulated annealing
  - TEMPI=20000: ΔE=27.5 kcal/mol
  - P=exp(- ΔE/kT)=0.5
  - So, high initial T will generate LOTS of structures
- Large amounts of output
  - Use print options to minimize this
  - Some structures (every NSTMIN steps) in .irc file
- How to avoid optimizing same structure many times?
  - Increase NSTMIN to 100 or more

Multi-Layered Approach to Solvation

EFP

Quantum Solute   Continuum (PCM)
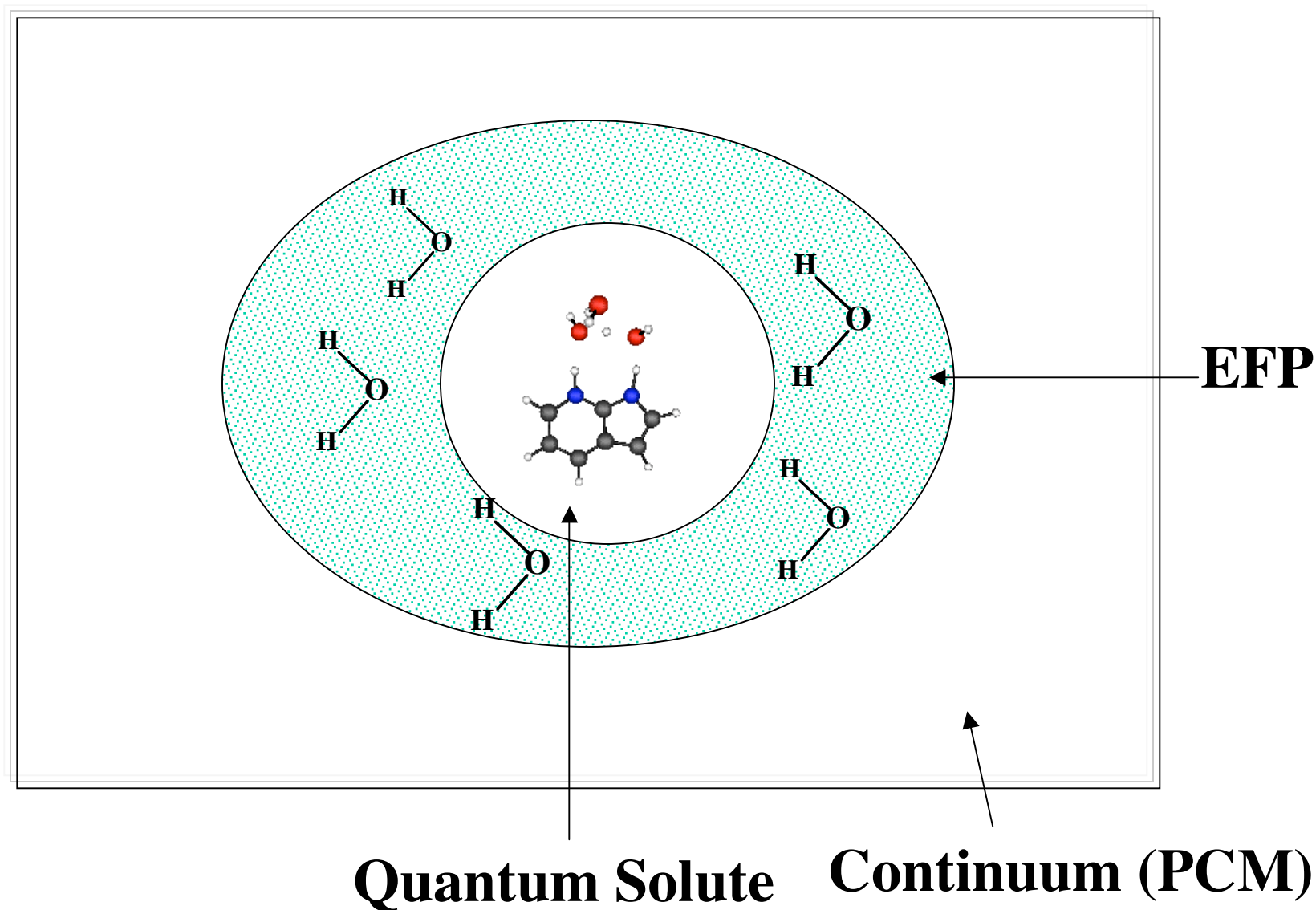
Interaction energy consists of : *electrostatic, polarization* and *exchange repulsion/charge transfer term*

$E_{interaction}$ = $E_{coulomb}$ + $E_{polarization}$ + $E_{exchange\ repulsion/charge\ transfer}$

$$E_{interaction} = \sum_{k=1}^{K} V_k^{Elec}(\mu, s) + \sum_{l=1}^{L} V_l^{Pol}(\mu, s) + \sum_{m=1}^{M} V_m^{Re\,p}(\mu, s)$$
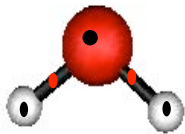
Distributed Multipolar expansion
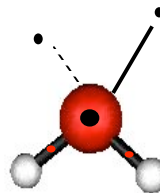
LMO polarizability expansion
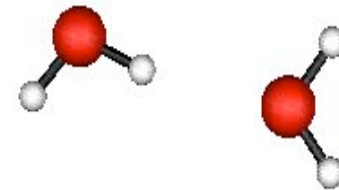
Fit to Functional Form

HIGHER-LEVELS OF EFP1

- DFT-based EFP (Ivana Adamovic)
  - Based on B3LYP: Same general approach
  - Adds some level of correlation

- MP2-based EFP (Jie Song)
  - Same general approach
  - Adds separate fit for dispersion
  - More effective correlation, especially long range
  - EFP-EFP done, EFP-QM in progress

GLYCINE: N vs. Z
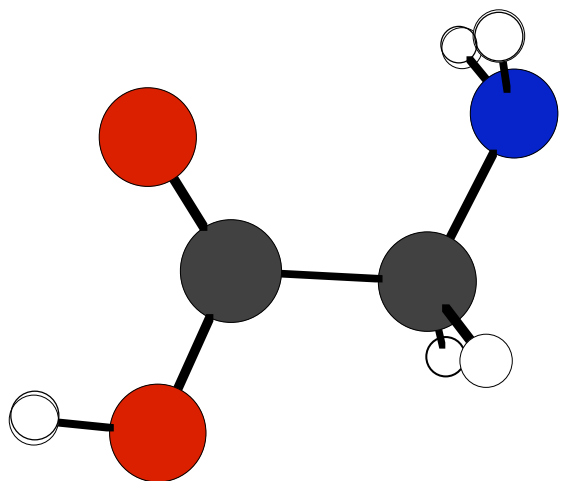
N
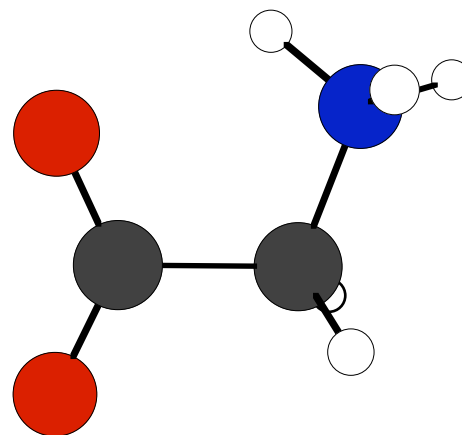
Z

Glycine: Christine Aikens
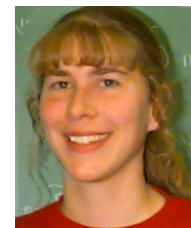
- Zwitterion (Z) in solution/crystal but nonionized (neutral) (N) in gas phase

- Experimental free energy/enthalpy difference in solution: 7.3/10.3 kcal/mol[1]

- Ionic complexes examined using mass spectrometric techniques

  - $[Gly(H_2O)_n]^-$ species with $n \geq 5$ seen; attributed to Z formation[2]

  - Smaller hydrates $Gly^- \cdot (H_2O)_{1-2}$ found later[3]

1. G. Wada, E. Tamura, M. Okina, M. Nakamura, *Bull. Chem. Soc. Jpn.* **55**, 3064 (1982).
2. S. Xu, J. M. Nilles, K. H. Bowen Jr., *J. Chem. Phys.* **119**, 10696 (2003).
3. E. G. Diken, N. I. Hammer, M. A. Johnson, *J. Chem. Phys.* **120**, 9899 (2004).

ISSUES

- At what size water cluster does Z become lower in energy than N?

- What is the converged N-Z energy difference?

- What are equilibrium structures for glycine-water complexes?

- What types of structures predominate for small clusters? larger clusters?

- How many waters are in the first solvation shell?

Previous Glycine Studies

- Many N gas phase conformers
- Lowest energy N conformer in continuum model differs from lowest energy N conformer in gas phase
- Z is not a local minimum in gas phase; minimum with PCM
- PCM+MP2 qualitatively correct (Z lower energy than N)
- Electrostatics stabilizes Z with respect to N
- Correlation energy stabilizes Z with respect to N

Previous Glycine Studies

- N: 1, 2 water molecules preferentially interact with COOH

- Z: 1, 2 water molecules bridge NH_3^+ and COO^- groups

- N lower in energy than Z for 1 or 2 waters

- Very few studies with more than two discrete water molecules

- Configurational sampling important, especially for larger numbers of waters
    - Mostly overlooked in previous studies

$\Rightarrow$ **Need for systematic study of the solvation of glycine**

Computational Methods (1-3 Waters)

- RHF/6-31G(d,p) optimizations for eight N conformers
- PCM+RHF/6-31G(d,p) optimization for Z glycine
- EFP2 potentials created for eight N minima, water, Z minimum, and three twisted Z structures
- Monte Carlo with local minimization ("basin-hopping") used with EFP2 potentials to find local minima
- Optimization with EFP1/HF potentials for water, RHF/6-31++G(d,p) for glycine
- Full optimization with RHF/6-31++G(d,p)
- Single point energies using MP2, PCM, PCM+MP2
- GAMESS used for all calculations

N Conformers

- 8 RHF/6-31G(d,p) gas phase N minima
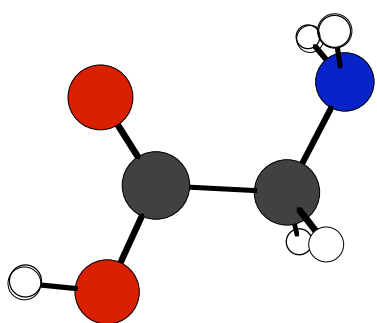
N1 0.0 0.0
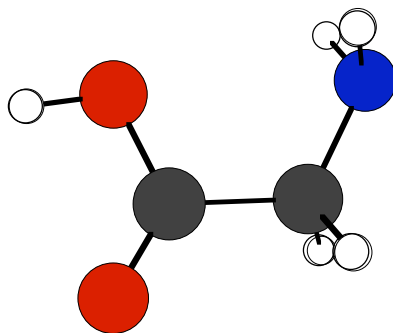
N2 1.9 1.6
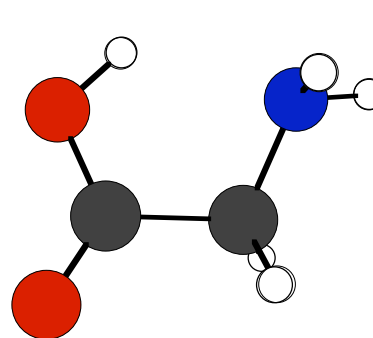
N3 3.0 1.0
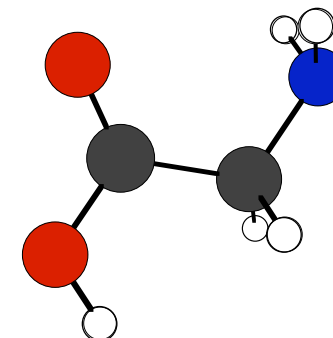
N4 6.9 6.1
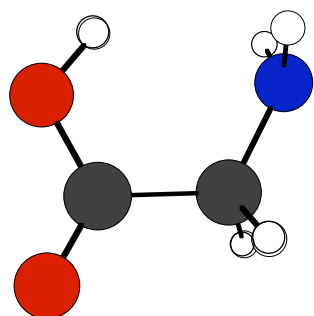
N5 9.0 7.3
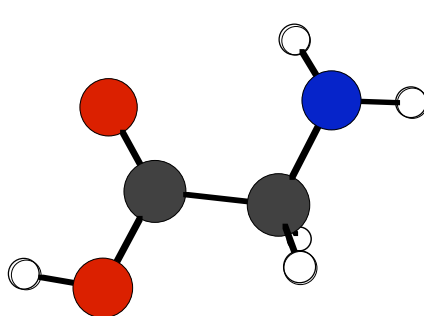
N6 2.0 1.4
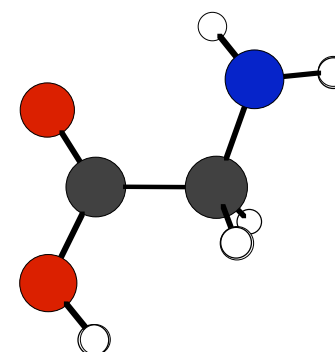
N7 9.2 7.7
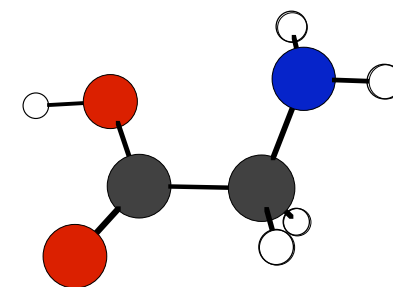
N8 3.1 2.3

RHF/6-31G(d,p)
MP2/6-31++G(d,p)//RHF/6-31G(d,p)

Z Structures

- No RHF/6-31G(d,p) Z minima
- One PCM+RHF/6-31G(d,p) Z
- Torsion around C-C, C-N bonds give Z′, Z″, Z‴

Z   Z′   Z″   Z‴
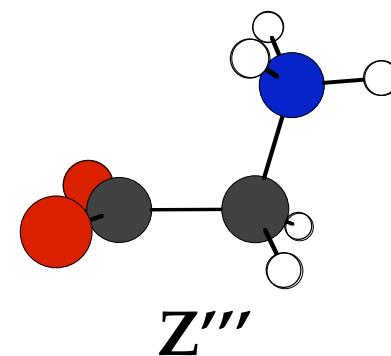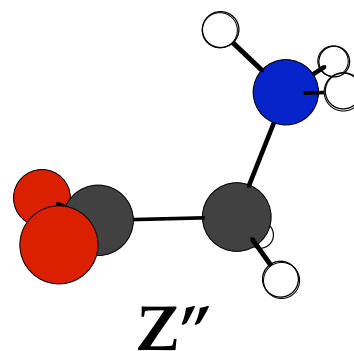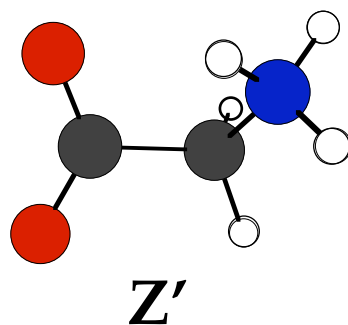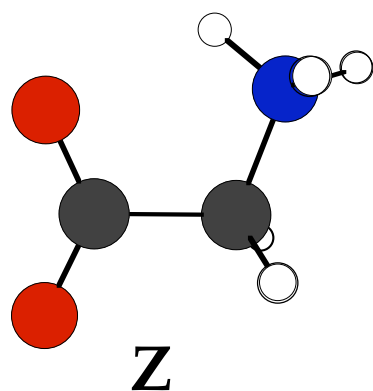
One Water

- 66 N, 25 Z structures from EFP2 Monte Carlo
- 47 N, 6 Z structures after mixed optimization
- 44 N, 5 Z structures after *ab initio* optimization
- More than one gas phase conformer yields low energy structures
- EFP1/HF waters track *ab initio* energies

Glycine(H_2O) N Structures

- Lowest energy N structures: water interacts with carboxylic acid
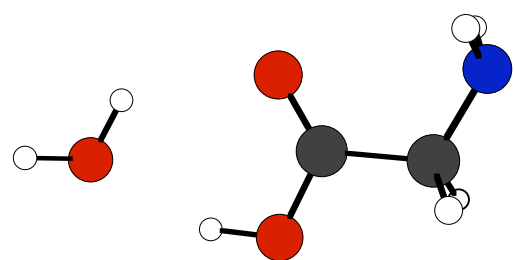
1N1-a 0.0 0.0 0.0 0.0 0.0
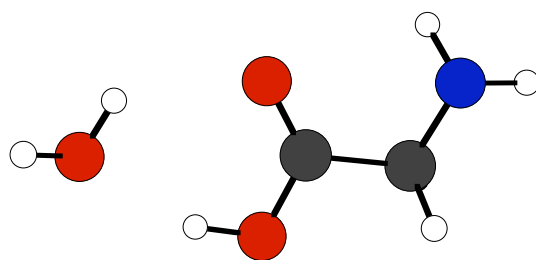
1N6-a 1.5 1.5 1.2 1.7 1.4
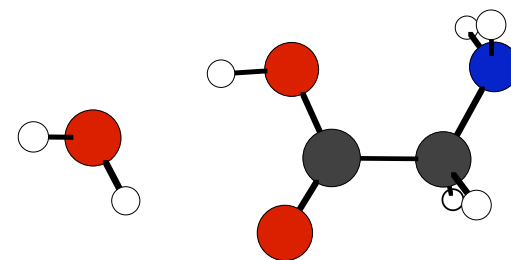
1N2-a 1.5 1.5 1.2 1.3 1.1
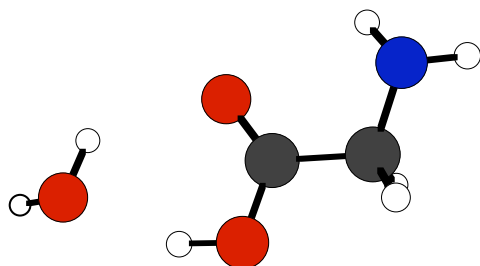
1N6-b 1.7 1.7 1.4 1.7 1.5
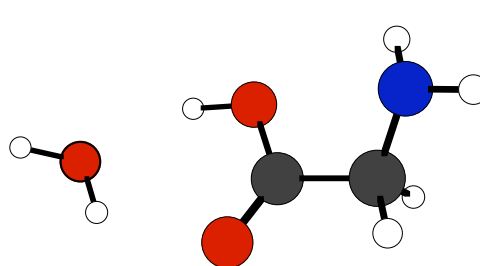
1N8-a 2.3 2.4 2.0 2.3 2.0
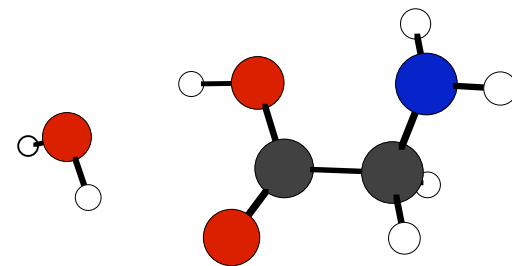
1N8-b 2.5 2.6 2.2 2.4 2.1

Glycine(H₂O) N Structures

EFP1/HF waters and RHF/6-31++G(d,p) glycine (kcal/mol)
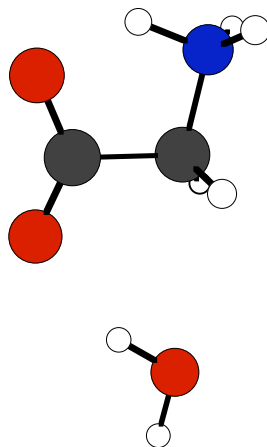RHF/6-31++G(d,p)
MP2//RHF/6-31++G(d,p)
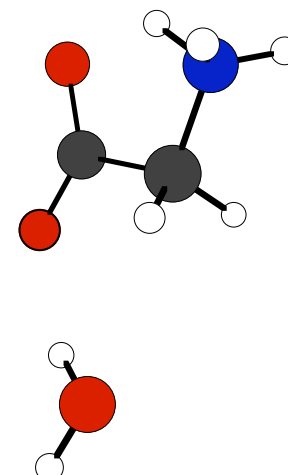PCM+RHF//RHF/6-31++G(d,p)
PCM+MP2//RHF/6-31++G(d,p)
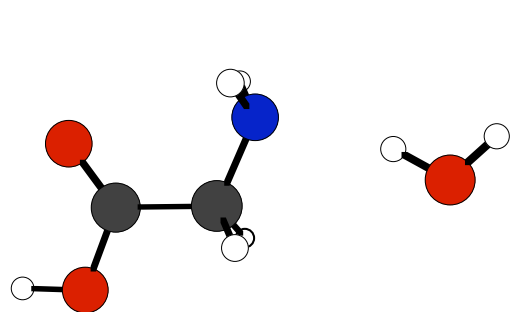
1N1-b 2.6 3.3 3.2 -0.2 0.1
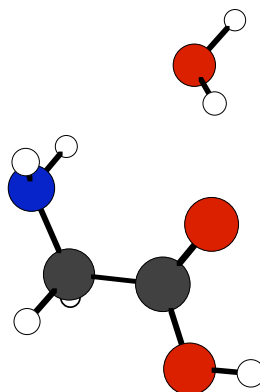
1N1-c 3.1 3.1 3.6 1.6 2.1
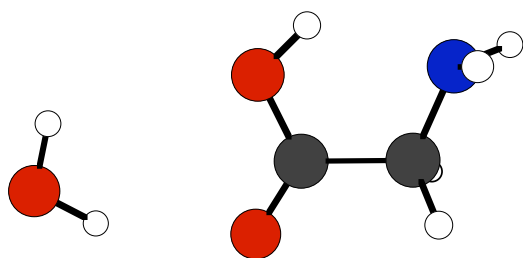
1N3-a 5.0 5.0 3.9 1.2 0.8

1N3-e 6.3 6.1 5.5 0.7 1.0
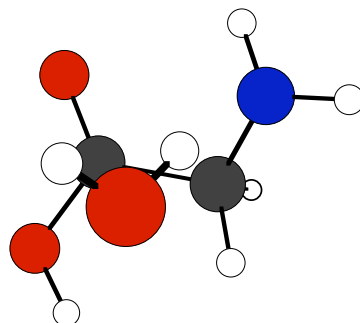
1N7-b 10.3 7.1 4.7 2.7 1.4
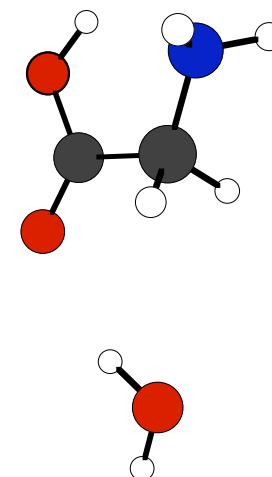
Glycine(H$_2$O) Z Structures

EFP1/HF waters and RHF/6-31++G(d,p) glycine (kcal/mol)
RHF/6-31++G(d,p)
MP2//RHF/6-31++G(d,p)
PCM+RHF//RHF/6-31++G(d,p)
PCM+MP2//RHF/6-31++G(d,p)

- Lowest energy Z structures: water bridges COO$^-$ and NH$_3^+$ groups
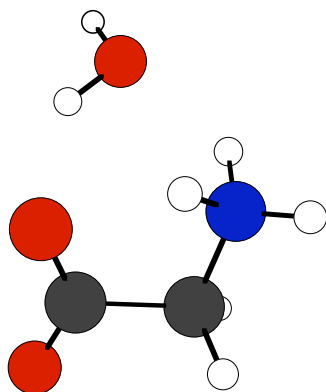
1Z-a 24.2 23.5 15.3 2.4 -3.6

1Z-c 27.2 26.7 17.5 4.8 -2.2

1Z-d 27.2 26.7 17.5 4.9 -2.2
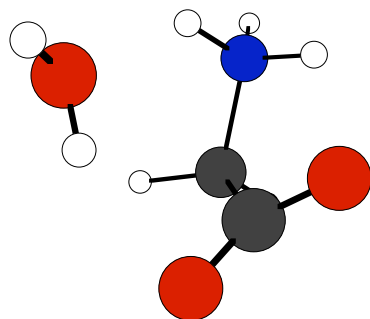
1Z-b 25.9 25.9 16.9 4.9 -2.0
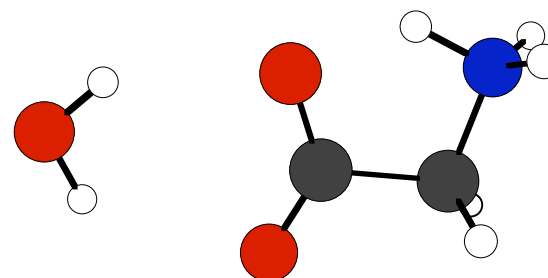
1Z-e 28.2 28.0 19.4 4.5 -1.7

Glycine(H$_2$O) Summary

- 1N1-a = lowest energy N structure for all methods
- 4 of the 6 lowest energy N structures not previously known
- N-Z energy difference: 24.2 kcal/mol with EFP1/HF waters
- MP2 single point energies stabilize Z with respect to N by 8.1-9.2 kcal/mol
- PCM single point energies stabilize Z by 21.0-23.5 kcal/mol
- **MP2+PCM predicts Z lower in energy than N by 3.6 kcal/mol**
- Z minima at RHF level
- MP2 optimization: Z becomes N

Two Waters

- 238 N, 86 Z structures from EFP2 Monte Carlo

- 155 N, 19 Z structures after mixed optimization

- 132 N, 11 Z structures after *ab initio* optimization

Glycine(H₂O)₂ N Structures

- **10 lowest energy N structures: waters form a ring with carboxylic acid end of glycine**

2N1-a 0.0 0.0 0.0 0.0 0.0

2N1-b 0.5 0.4 0.6 0.3 0.6

2N2-a 1.4 1.3 1.1 1.3 1.1

2N6-a 1.5 1.5 1.2 1.6 1.3
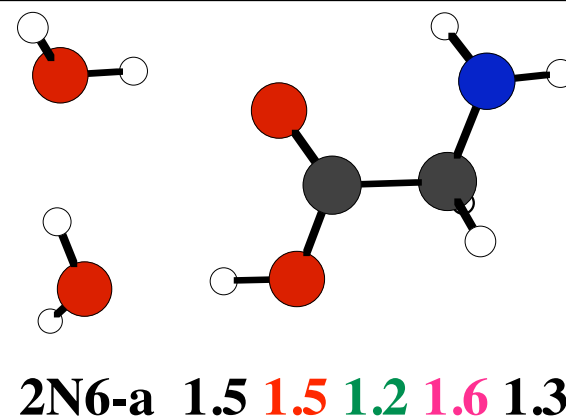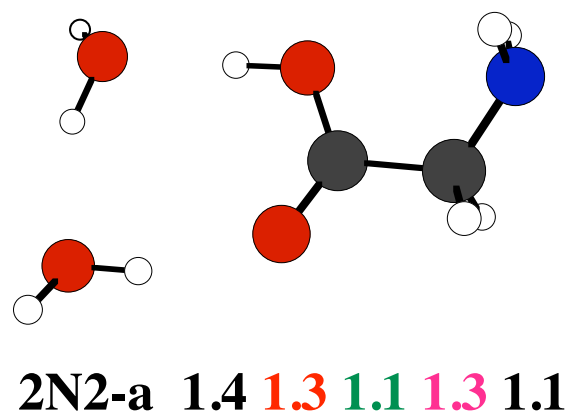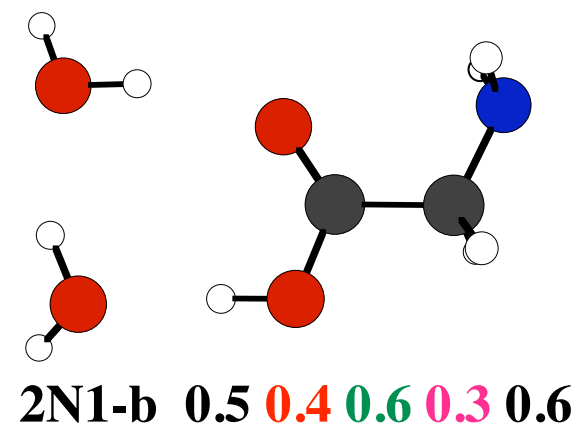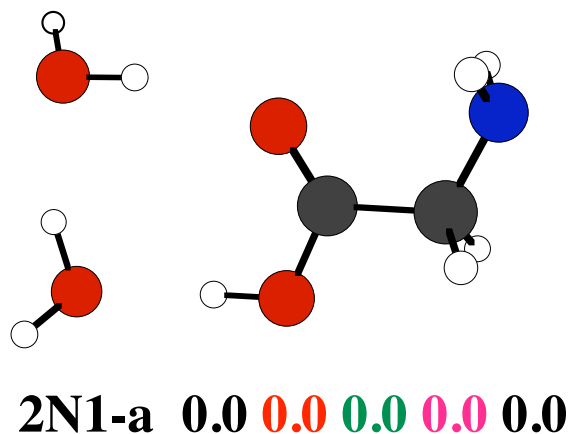
Glycine(H_2O)$_2$ Z Structures

EFP1/HF waters and RHF/6-31++G(d,p) glycine (kcal/mol)
RHF/6-31++G(d,p)
MP2//RHF/6-31++G(d,p)
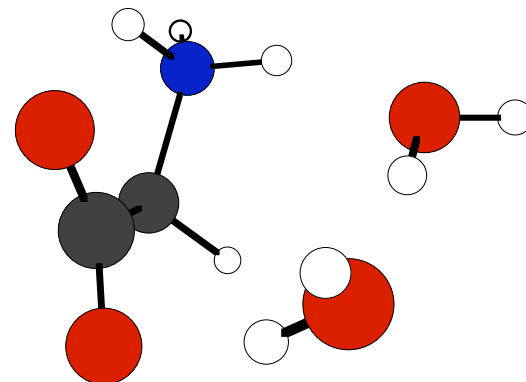PCM+RHF//RHF/6-31++G(d,p)
PCM+MP2//RHF/6-31++G(d,p)

- Lowest energy Z structures: waters bridge COO$^-$ and NH$_3^+$ groups
  - **Two water bridge**
  - Two individual bridges
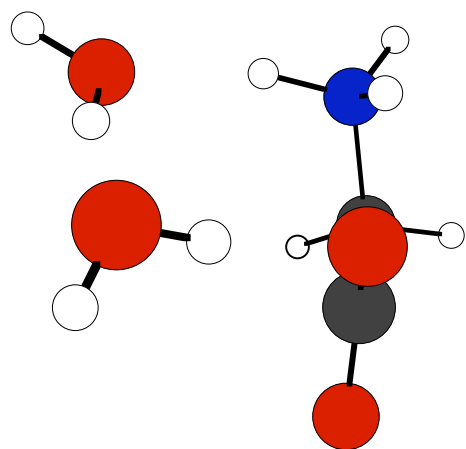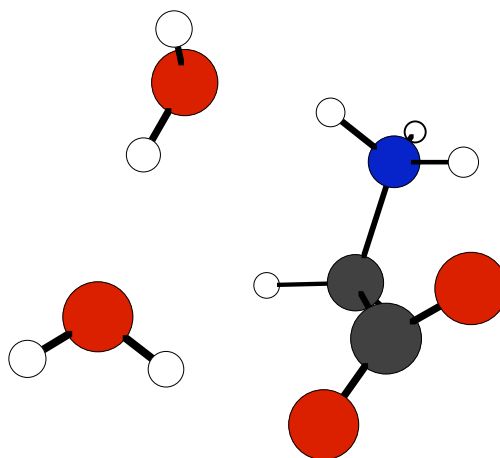
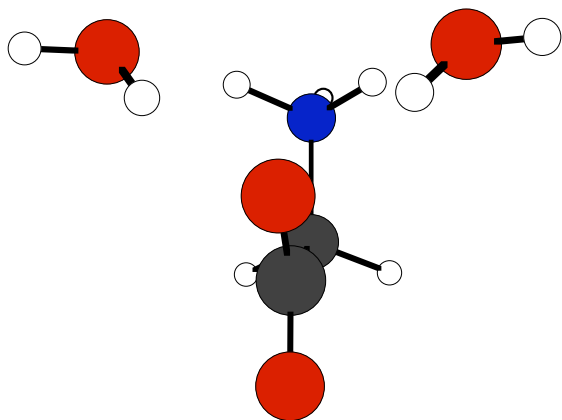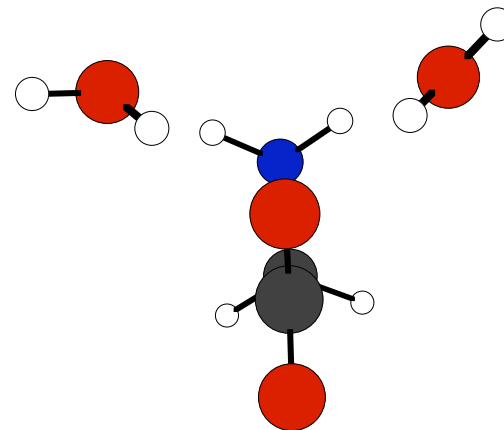2Z-a 20.2 19.8 10.8 2.4 -4.7 | 2Z-b 20.6 20.6 11.3 3.0 -4.4 | 2Z-c 20.9 20.6 11.4 3.4 -3.9

Glycine(H₂O)₂ Z Structures

- Lowest energy Z structures: waters bridge COO^- and NH_3^+ groups
  - Two water bridge
  - **Two individual bridges**

2Z-d 21.3 19.7 11.4 2.7 -3.7

2Z-e 21.3 19.7 11.4 2.8 -3.7

Glycine(H$_2$O)$_2$ Summary

- N-Z energy difference: 20.2 kcal/mol with EFP1/HF waters
- EFP1/HF waters track *ab initio* calculations closely
- MP2 stabilizes Z relative to N by 7.0-9.4 kcal/mol
- PCM stabilizes Z by 16.0-23.1 kcal/mol
- **PCM+MP2 predicts Z more stable than N by 4.7 kcal/mol**
- MP2 optimizations: 8 unique Z structures have bridging water molecule(s)

Three Waters

- 824 N, 212 Z structures from Monte Carlo
- 450 N, 42 Z structures after mixed optimization
- 349 N, 24 Z structures after *ab initio* optimization (24 N within 3.0 kcal/mol)
- Lowest energy N structures:
  - Water ring at carboxylic acid end of glycine
  - Bridge between carboxylic acid group and nitrogen atom
- Lowest energy Z structures: waters bridge COO^- and NH_3^+ groups
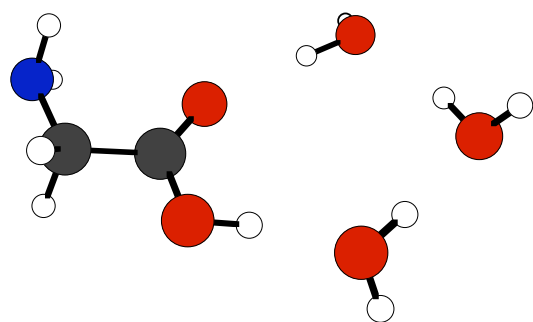
Glycine(H$_2$O)$_3$ Structures

EFP1/HF waters and RHF/6-31++G(d,p) glycine (kcal/mol)
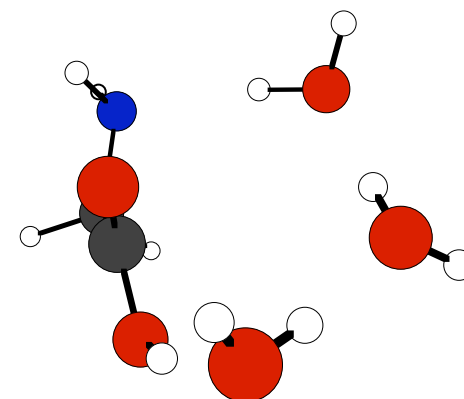RHF/6-31++G(d,p)
MP2//RHF/6-31++G(d,p)
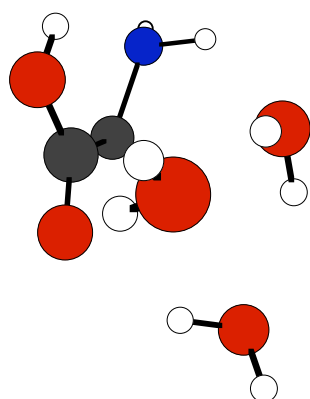PCM+RHF//RHF/6-31++G(d,p)
PCM+MP2//RHF/6-31++G(d,p)

3N1-a 0.0 0.0 0.0 0.0 0.0
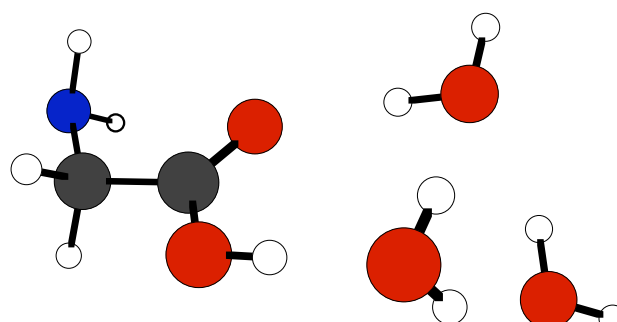
3N1-b 0.4 1.3 1.0 1.9 1.6

3N6-a 1.1 2.9 0.2 2.3 0.1

3N3-h 5.5 6.5 3.9 6.7 4.3
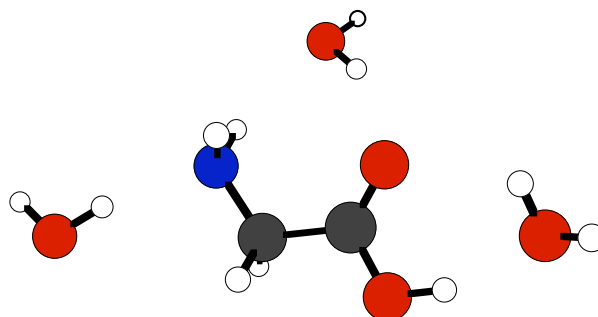
3N1-ar 6.6 7.1 7.3 4.1 4.2
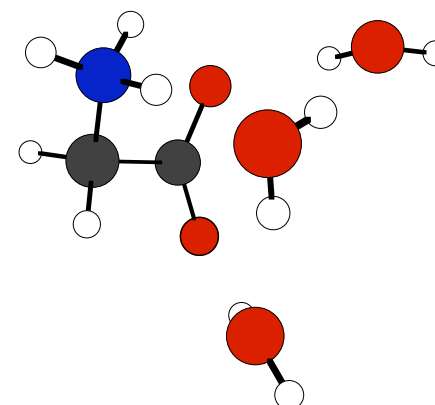
3Z-a 14.7 15.6 5.2 1.9 -6.9

Glycine(H$_2$O)$_3$ Summary

- N-Z energy difference: 14.7 kcal/mol with EFP1/HF waters

- MP2 stabilizes N bridge structures relative to carboxylic acid ring structures

- MP2 stabilizes Z relative to N by 6.7-10.4 kcal/mol

- PCM stabilizes Z with respect to N by 12.1-18.7 kcal/mol

- **PCM+MP2 predicts Z to be 6.9 kcal/mol lower in energy than N**

Computational Methods
(4+ Water Molecules)

- EFP2 Simulated annealing Monte Carlo with local minimization to find minima
- EFP1/HF potential for waters
- RHF/6-31++G(d,p) for glycine
- Starting structures from mixed optimizations with 3 waters (450 N, 42 Z)
- Full optimizations using
  RHF/6-31++G(d,p) for low energy structures
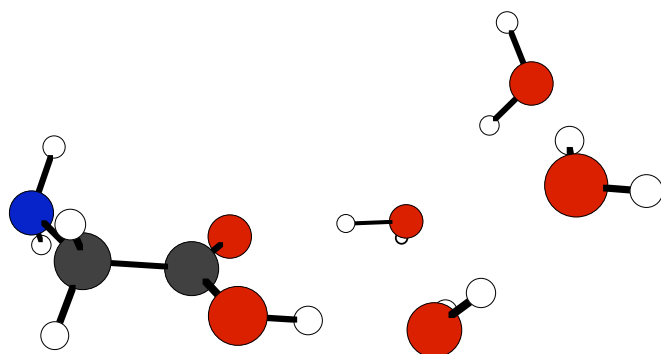- Single point energies using MP2, PCM, PCM+MP2

Glycine(H$_2$O)$_4$ Structures

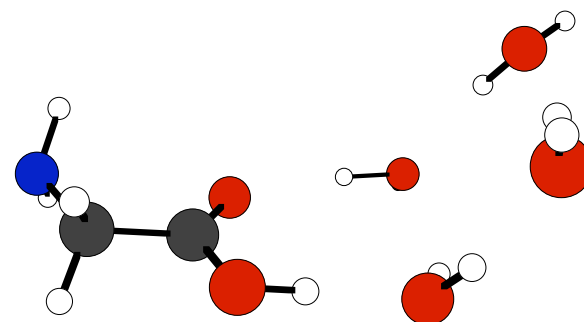EFP1/HF waters and RHF/6-31++G(d,p) glycine (kcal/mol)
RHF/6-31++G(d,p)
MP2//RHF/6-31++G(d,p)
PCM+RHF//RHF/6-31++G(d,p)
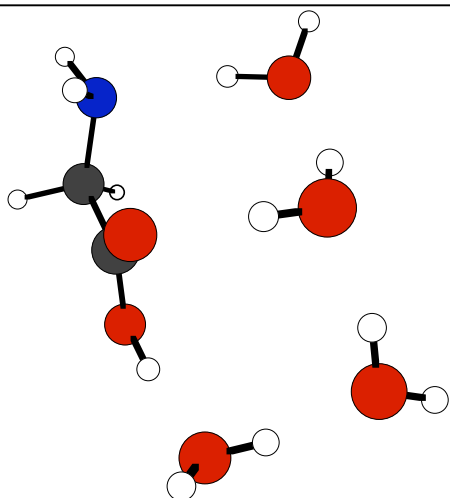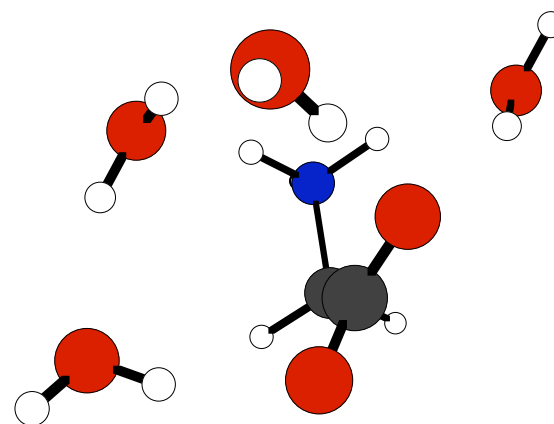PCM+MP2//RHF/6-31++G(d,p)

4N1-a 0.0 0.0 0.0 0.0 0.0

4N1-b 0.3 0.4 0.5 0.0 0.1

4N6-a 0.3 0.8 -1.0 0.4 -1.0

4Z-a 12.1 11.7 1.6 1.2 -7.7

Glycine(H$_2$O)$_4$ Summary

- 26 N structures from N1 within 3.0 kcal/mol
- 43 Z structures within 3.0 kcal/mol
- N6 structure ("more solvated")
  - 1.7 kcal/mol more stable after MP2 optimization
- N-Z energy difference is 12.1 kcal/mol with EFP1/HF waters
- MP2 stabilizes Z structures by 8.5-11.7 kcal/mol
  - N still lower in energy
- PCM stabilizes Z structures by 10.0-14.7 kcal/mol
- **PCM+MP2 predicts 4Z-a to be:**
  - 7.7 kcal/mol lower in energy than 4N1-a
  - 6.7 kcal/mol lower in energy than 4N6-a

Five Waters

- Lowest energy structure comes from N6
- Lowest energy structure from N1 not very solvated
- N-Z energy difference is 8.6 kcal/mol with EFP1/HF waters
- MP2 stabilizes Z structures by about 9 kcal/mol
- PCM stabilizes Z structures by about 9 kcal/mol
- **PCM+MP2 predicts Z to be lower in energy than N by 7.3 kcal/mol**
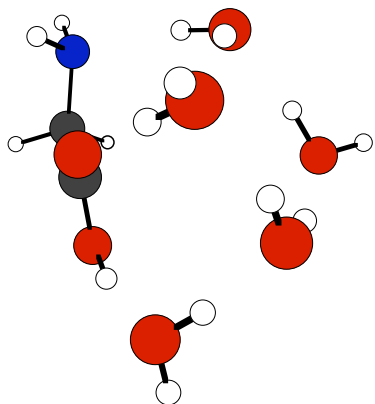
Glycine(H$_2$O)$_5$ Structures

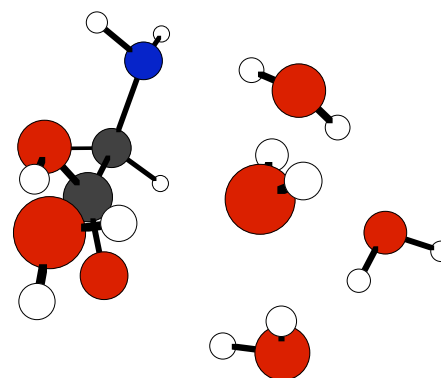EFP1/HF waters and RHF/6-31++G(d,p) glycine (kcal/mol)
RHF/6-31++G(d,p)
MP2//RHF/6-31++G(d,p)
PCM+RHF//RHF/6-31++G(d,p)
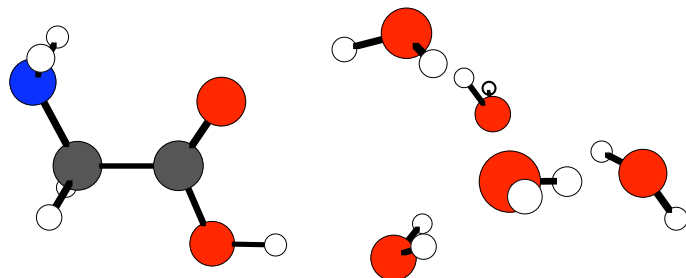PCM+MP2//RHF/6-31++G(d,p)

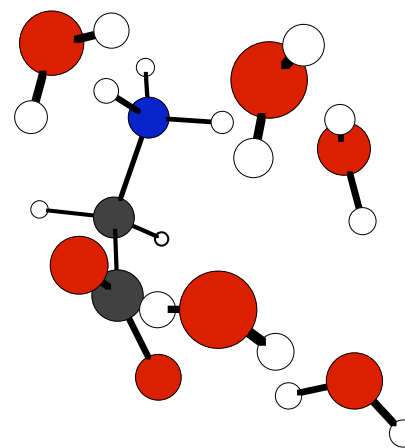**5N6-a 0.0 0.0 0.0 0.0 0.0**

5N8-a 0.3 1.0 0.3 1.5 0.8

5N1-a 0.4 0.2 2.2 -0.1 1.7

5Z-a 8.6 9.5 0.4 0.8 -7.3

Glycine(H$_2$O)$_6$

- N-Z energy difference = 8.4 kcal/mol with EFP1/HF waters

- PCM stabilizes Z by 8.6 kcal/mol

- MP2 stabilizes Z by 7.6 kcal/mol

- **PCM+MP2 predicts Z more stable by 6.9 kcal/mol**

6N6-a 0.0 0.0 0.0 0.0 0.0   6N1-a 0.6 0.3 1.5 0.1 1.0   6Z-a 7.4 8.4 0.8 -0.2 -6.9
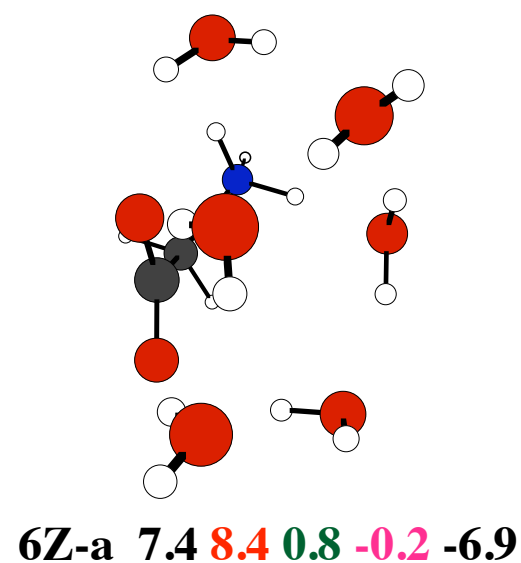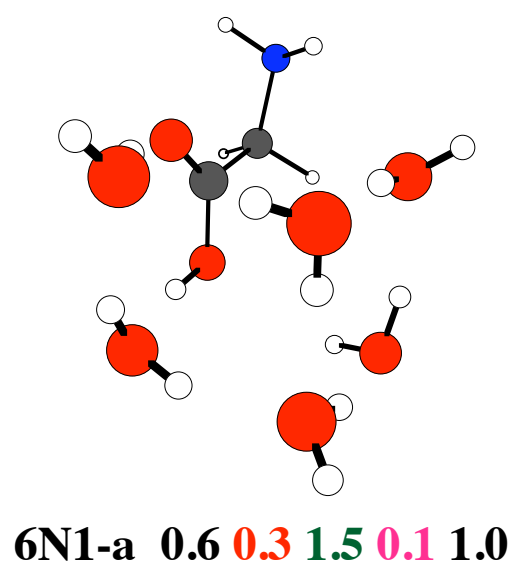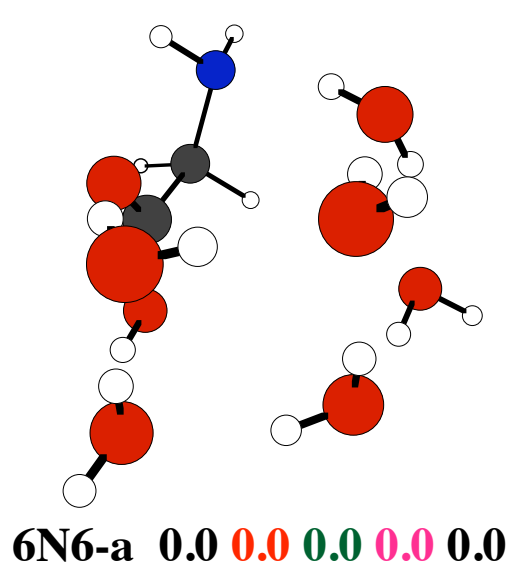
Seven Waters

- MP2: 7N8-a lowest energy N structure
- N-Z energy difference is 7.3 kcal/mol with EFP1/HF waters
- PCM stabilizes Z structures by ~ 8 kcal/mol
- MP2 stabilizes Z structures by ~ 9 kcal/mol
  – 7Z-b predicted to be global minimum
- **PCM+MP2: Z more stable by 8.8 kcal/mol**

Glycine(H₂O)₇ Structures

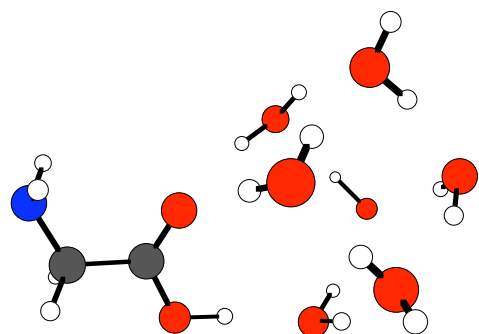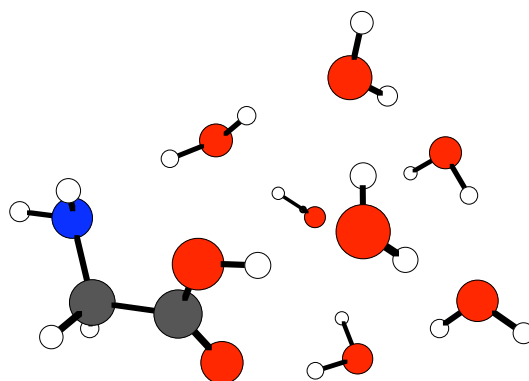EFP1/HF waters and RHF/6-31++G(d,p) glycine (kcal/mol)
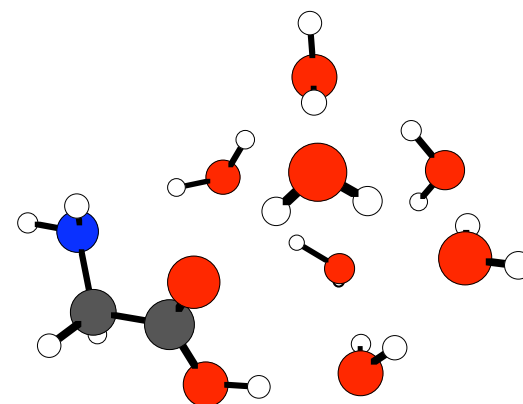RHF/6-31++G(d,p)
MP2//RHF/6-31++G(d,p)
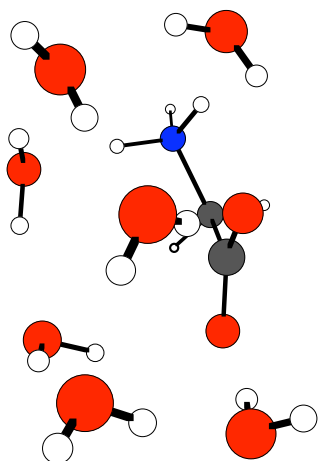PCM+RHF//RHF/6-31++G(d,p)
PCM+MP2//RHF/6-31++G(d,p)

7N1-a 0.0 0.0 0.0 0.0 0.0

7N8-a 0.5 1.2 -1.0 1.6 0.0

7N6-a 0.7 1.3 -0.7 2.1 0.6

7Z-a 7.3 9.0 -0.7 0.7 -7.8
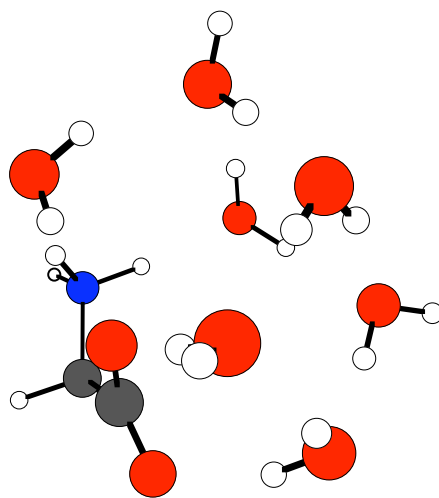
7Z-b 7.4 8.5 -1.3 -0.5 -8.8
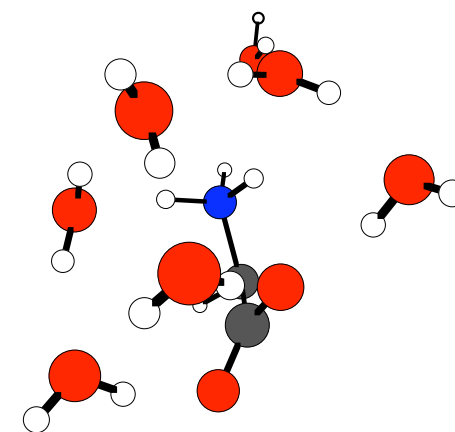
7Z-f 8.9 8.4 -0.5 2.5 -5.4

Glycine Conclusions (So Far)

- *Many* energy structures
- Configurational sampling important!
- Multiple conformers must be considered
- EFP1/HF potentials track RHF structures and energies well
- Correlation energy stabilizes zwitterion (~8-9 kcal/mol)
- Continuum solvation stabilizes zwitterion; effect decreases with more discrete waters
- Without PCM, N lower in energy than Z for 6 or fewer waters
- MP2: Z with 7 waters lower in energy than N
- **PCM+MP2: Z structure ~7 kcal/mol lower in energy than N (approximate convergence after 3 waters)**

WHAT'S NEXT?

- Increase # waters

- Determine convergence of N-Z energy difference

- Determine size of first solvation shell and # waters needed to fully solvate glycine